Distributed intelligent data analysis in diabetic patient management

R. Bellazzi, C. Larizza, A. Riva, A. Mira, S. Fiocchi* and M. Stefanelli
Dipartimento di Informatica e Sistemistica, Università di Pavia, Pavia, Italy
* I.R.C.C.S. Policlinico S. Matteo, Pavia, Italy

This paper outlines the methodologies that can be used to perform an intelligent analysis of diabetic patients' data, realized in a distributed management context. We present a decision-support system architecture based on two modules, a Patient Unit and a Medical Unit, connected by telecommunication services. We stress the necessity to resort to temporal abstraction techniques, combined with time series analysis, in order to provide useful advice to patients; finally, we outline how data analysis and interpretation can be cooperatively performed by the two modules.

INTRODUCTION

The crucial role of Intensive Diabetes Therapy (IDT) in delaying or preventing the development of long-term complications of Insulin Dependent Diabetes Mellitus (IDDM) has been recognized in several recent studies [1]. On the other hand, the negative implications of IDT implementation in daily patients management basically consist in increasing the risk of severe hypoglycemia and in the potential increase of the costs of the therapy, due to the continuous assistance required. These two negative features could be avoided by exploiting current advances of information technologies, particularly telecommunications networks and knowledge-based systems. Like other authors [2, 3], we believe that effectiveness and safety of IDT could be increased if the rate of information transmission between patients and clinicians were increased; moreover, we believe that patients should be involved in their own therapy management, through a careful home assistance provided by a cooperation between patients and physicians. Such cooperation can be effectively implemented in a distributed environment for patient care [4].

The above mentioned motivations and previous experiences [2], led to the definition of the T-IDDM *(Telematic Management of Insulin-Dependent Diabetes Mellitus)* project[4, 5, 6].

In the T-IDDM architecture the IDDM patient management is divided into several subtasks, that are implemented by two basic components: a *Patient Unit* (PU) and a *Medical Unit* (MU), interconnected by a telecommunication system. The MU is designed to assists the physician in defining a *treatment protocol*, by suggesting insulin regimen, diet and physical exercise, through a periodic evaluation of patient's data. The treatment protocol is then communicated to the PU in order to bind the space of its admissible actions. The PU assists the patients in their self-monitoring activity, by giving proper therapeutic advice, such as insulin dosage adjustments. Moreover, the PU deals with automatic data collection and transmission from the patient's house to the clinic, by communicating the therapeutic actions and the current metabolic state. The telecommunication system relies on an HTTP server, developed in Common Lisp; hence, the integration of the reasoning tools that constitute the MU is based on the HTTP protocol, while user interaction takes place using the HTML language [7].

In this paper we will describe the data interpretation and plan revision tasks performed by the two modules and we will show how the cooperation and coordination between the modules can be useful to efficiently and effectively assist patients and physicians in making decisions.

THE PROBLEM

In order to understand the complexity of IDDM patient's data analysis let us consider Fig. 1. The displayed Blood Glucose Level (BGL) time series comes from the home monitoring of a 20 year old male IDDM patient. The patient underwent three insulin injections per day, at breakfast, lunch and dinner time, respectively. The insulin injections at breakfast and dinner were a mixture of regular and NPH (intermediate) insulin. BGL measurements are performed before meals, as required by the Diabetologic Unit of the Pavia University Medical School. The mean

0195—4210/96/$5.00

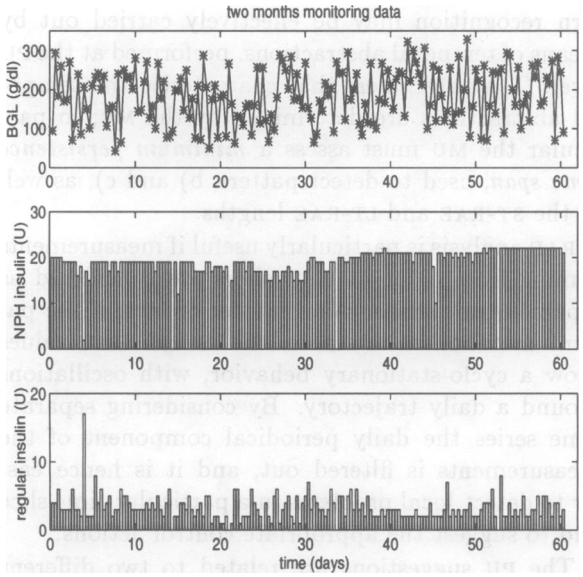

Figure 1: BGL values and insulin doses over a period 60 days (see text).

of BGL measurements over two months was 182.7 mg/dl while the standard deviation was 60 mg/dl. Only a nominal diet plan was known, and there is no assurance that it was actually followed on each day. The patient gave only qualitative information on the physical exercise performed. The time series is highly unstable and on the basis of the available data it was neither possible to derive significant causal models nor to obtain reliable estimates of the parameters of mathematical models describing BGL or insulin dynamics. In clinical practice, this kind of situation is unfortunately very frequent; in order to determine whether a computerized distributed system is able to provide useful advice to patients and physicians in such a context, it is hence necessary to answer two fundamental questions: What kind of information do physicians need from a system able to automatically analyze this kind of data? How it is possible to use such low quality information to help patients in properly and safely modifying their therapy?

METHODS

From a functional point of view, the two units of T-IDDM architecture (PU and MU) cooperate to perform a hierarchical adaptive control system of the metabolic state of IDDM patients [5]. We will now describe a complete control cycle, from the assessment of the therapeutic protocol to its revision in response to changes in the patient's status, in order to highlight what methodologies are used to analyze the data, and how their results are exploited by the two modules.

Protocol assessment

In current medical practice, the patient diary is revised and a new therapeutic protocol is assessed only when a patient undergoes the periodic control visit. On the contrary, the MU is able to dynamically suggest to the physicians a protocol updating, by combining the home monitoring data coming from the PU with the historical information stored in the patient data base. The MU knowledge-based system relies on an ontology of IDDM and on a number of inference mechanisms [6] to perform a three step procedure: a *data interpretation* task extracts high-level metabolic and statistical parameters from the individual measurements; a *reasoning* task applies a logic-based belief-maintenance algorithm in order to evaluate the state of the patient and a *decision* task exploits the results of the first two tasks to choose or to adjust a protocol using heuristic or model-based techniques.

In order to allow a proper interpretation of the data, the MU subdivides the 24-hour daily period into a set of consecutive non-overlapping time slices. These time slices are generated on the basis of the information about the patient's life style, in particular the meal times. The possible adjustments are then selected using the concept of *competent time slice*: an action in a certain time slice will be competent for the BGL measurements in the time slices that it directly affects. For example, an intake of regular insulin will be competent for the time slices that cover the subsequent six hours. Therefore, when a problem is detected in a particular time slice t, the possible adjustments will be the ones affecting the actions in the time slices that are competent over t.

This information is used by the MU to suggest a new therapeutic protocol, that is passed to the PU. The protocol is composed of suggested actions (insulin intakes, diet, exercise) with their competent time slices, and of the PU control tables, that specify the strategies for coping with dangerous situations in the different time slices [6].

Patient advice

The PU has two fundamental requirements: it must react to dangerous events in a ***fast and proper way*** and it must be able to provide suggestions relying on ***limited computational capabilities***. The reasoning procedures that it uses to give therapeutic suggestions are guided and pre-determined by the MU through the control tables transmitted along with the therapeutic protocol. A definition of control tables, based on meal intakes and measured and predicted BGLs, has been previously proposed by the au-

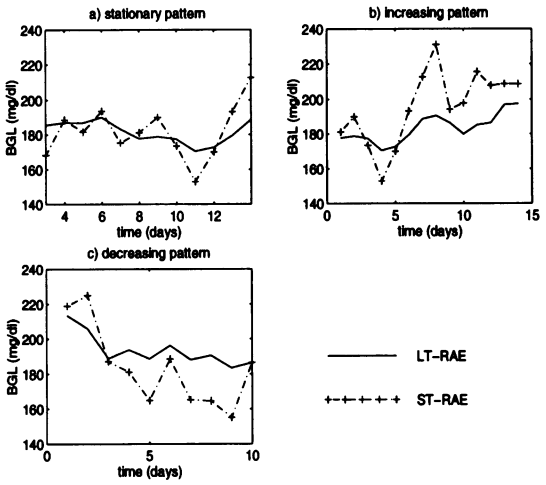

Figure 2: The three patterns derived from RAE analysis: a) stationary pattern; b) increasing pattern; c) decreasing pattern

thors [5]; this approach is able to explicitly take into account inter- and intra-individual variability. Nevertheless, when the self-monitoring data contain only BGL and glycosuria values, while meals and physical exercise are only approximatively known, the control tables should be based on more robust techniques [8]; in particular, we are currently designing control tables that rely on a cooperative data analysis between the PU and the MU.

At the PU level, we are now considering a simple and efficient technique that exploits the *Running Average* calculation to detect stable trends in BGL time series. Given a generic unidimensional time series, the basic Running or Moving Average Estimate (RAE) for an observation is computed by calculating the average over the k preceding values; k is called the *running average length*. Modifications of the basic technique able to take into account linear trends in time can be found in [9]. This technique is useful to detect local trend components in the BGL time series, exploiting two different RAEs. The first RAE (long-time RAE, LT-RAE) has a large value of k, and therefore provides smooth estimates of the local trend; the second RAE (short-time RAE, ST-RAE) has a smaller value of the RA length, and is hence more sensitive to the BGL deviation from the overall mean.

The relationships between the two curves identify three different patterns (see Fig. 2): a stationary pattern a), an increasing trend pattern b) and a decreasing trend pattern c). Patterns b) and c) show an abrupt change in the ST-RAE derivative, and are characterized by a crossing point after which the ST-RAE is *persistently* higher or lower, respectively, than the LT-RAE. When applied to BGL time series, pattern recognition may be effectively carried out by means of temporal abstractions, performed at the PU level. The parameters that characterize the temporal abstractions are determined by the MU; in particular the MU must assess a *minimum persistence time span*, used to detect pattern b) and c), as well as the ST-RAE and LT-RAE lengths.

RAE analysis is particularly useful if measurements coming from different time slices are considered as separate time series. As a matter of fact, if the patient metabolism is stable, the overall BGL values show a cyclo-stationary behavior, with oscillations around a daily trajectory. By considering separate time series the daily periodical component of the measurements is filtered out, and it is hence easier to detect local problems in a particular time slice and to suggest the appropriate control actions.

The PU suggestions are related to two different therapy adjustments. The first one is based on the present BGL value, and attempts to react against dangerous instantaneous situations (for example, by modifying a single regular insulin dose or delaying a meal). The second therapy adjustment is based on the above described detection of local BGL trends and on a set of control tables. Each table is based on three inputs: the ST-RAE, the current ST- and LT-RAE patterns and the glycosuria level. The table output is a control action that deals with non-stationary behaviors by suggesting a *permanent* modification to the most competent protocol action for the identified problem (for example, changing an NPH insulin dose). A different control table is generated by the MU for each time slice, according to the constraints imposed by patient's life styles.

It is important to notice that the application of a control action of the second type may be delayed until the MU confirms it. This can happen when the PU generates advice that may lead to hypoglycemic episodes; for example, when the PU suggests an insulin increase due to an increasing BGL pattern, but isolated hypoglycemic events have also been observed. Moreover, before modifying the protocol, the MU may also require additional laboratory parameters, like the HbA1C value.

Figures 3 and 4 show an example of the application of the RAE method to the data of a single time slice. In this case we analyzed the series of Before Breakfast BGL (BB-BGL) measurements of the data displayed in Fig. 1.

The BGL values increase significantly after day 30, although the variance is still very high. The ST-RAE is calculated with an RA length of 7 days, and the LT-RAE is calculated with an RA length of 21 days; the two curves allow for a straightforward interpretation

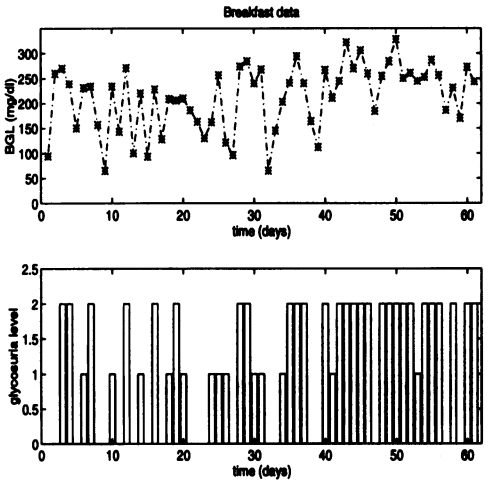

Figure 3: BB-BGL values over 60 days. An increasing trend after day 30 is clearly identifiable

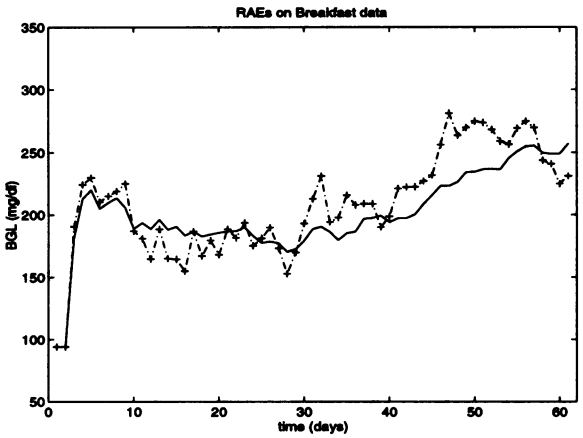

Figure 4: RAEs with running average length of 7 days (dotted line) and 21 days (continuous line)

of the different local patterns in the 60 days follow-up period. If patterns b) and c) are identified with a persistence of at least 4 days without intersection between the two curves, we detect two non-stationary patterns: the first one, of type c), from day 10 to 22, and the second one, of type b), from day 30 to 56.

This technique, used for retrospective analysis of 6 patients monitored over 6 months, detected 98 instances of non-stationary patterns, 95 of which were judged to be clinically significant. The relevant episodes were detected in advance with respect to visual inspection by the physician, and the results obtained were often easier to interpret than the ones derived with more sophisticated techniques [5]. We are now evaluating the RAE technique in comparison with other low-computational-cost methods, such as piecewise linear autoregressive models.

Protocol revision

The PU and the MU work asynchronously, since the PU is not continuously connected with the MU, and, although periodical communications are required, it is not a-priori known when the PU will transmit the data to the MU. Whenever a dangerous condition, such as an episode of non-stationarity, is detected, the PU may suggests protocol revisions: the modified protocol must nevertheless be checked and confirmed by the MU. When a new connection is established, the PU sends the data analysis results to the MU, together with the monitoring data and the suggested actions. The MU will check the adequacy of the actions by applying a number of available *data abstraction* methods.

In particular, Temporal Abstractions (TA) have been recognized to be of fundamental importance in order to help data interpretation [10]. The TA mechanisms adopted at MU level have been described in a different application context [11]. Following the TA ontology used in [11] and according to the domain medical knowledge, we defined a set of relevant critical situations that may be efficiently recognized through TA. A subset of the TAs defined for the breakfast time-slice is shown in the following table:

TA type	Temporal Abstractions
state	hypoglicemia, hyperglycemia, glycosuria, extra physical exercise
trend	BGL increase, BGL decrease, BGL stationarity
complex	Somogy effect, Dawn effect, improper patient action

The complex TAs reported in the table are used to detect critical situations, represented by different combinations of glycosuria, BGL, and life-style indicator values. *Somogy effect*, defined as a response to hypoglycemia while asleep with counter regulatory hormones causing morning hyperglycemia, is detected by looking for "hyperglycemia at breakfast with absence of glycosuria"; *Dawn effect*, a morning hyperglycemia unrelated to nightly hypoglycemia, is detected by searching for "hyperglycemia at breakfast with presence of glycosuria", and finally, *Improper Patient Action* refers to hypoglycemia or hyperglycemia episodes caused by a wrong protocol implementation (e.g. an unpredictable change of the life style or an insulin dose delivery error).

The derived TAs are exploited by the reasoning module of the MU to confirm or reject the protocol adjustments suggested by the PU. After being approved and possibly modified by physicians, the inference result is transmitted back to the PU in the form of a revised protocol.

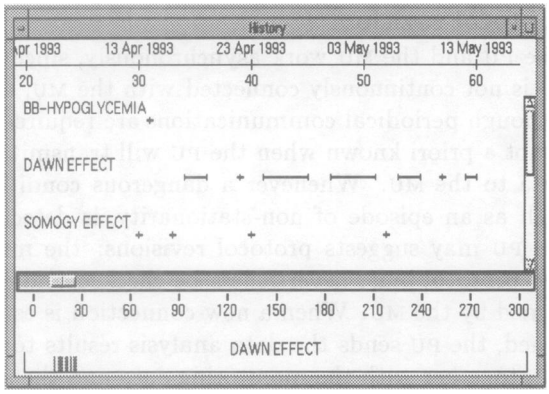

Figure 5: Temporal abstractions for the breakfast time slice

In the example described above, after day 30 the PU would suggest an increase of the insulin dose competent for the Breakfast Time Slice, i.e. NPH insulin delivered at dinner. Nevertheless, the occurrence of a hypoglycemic episode at day 32 would force the PU to suggest a life-style check while waiting for the MU modification of the therapeutic protocol. By analyzing the data from day 30 to 45, the MU could detect a persistent Dawn effect following the hypoglycemic episode, together with isolated cases of Somogy effect (see Fig. 5). Instead of accepting the PU suggestion, the MU would therefore propose a protocol modification, in which the NPH injection is delayed to bed time to reduce the risk of night hypoglycemia, and submit it to the physician for approval.

DISCUSSION AND FURTHER RESEARCH

In this paper we have shown a part of the methodologies that will be involved in the definition of a system for telematic management of IDDM patients. We addressed two important problems related to the analysis of patient data and their use for therapeutic purposes: we showed that a system able to automatically perform time series analysis and temporal abstractions can provide useful support to physicians also in the presence of low-quality data, and that the cooperation between two distinct control modules may help patients in properly and safely modifying their therapy. The ideas here presented will be fully integrated in a prototypical system that is being implemented within the T-IDDM EU project, and will be evaluated during the demonstration phase of the project.

Acknowledgments

This work is part of the T-IDDM project HC-1047, funded by the European Commission in the Telematics Application Programme.

References

1. The Diabetes Control and Complication Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus, *The New England Journal of Medicine*, 1993: 14-329: 977-986.
2. Gomez EJ, Del Pozo F, Arredondo MT, Sanz M, Hernando E. A Telemedicine Distributed Decision-Support System for Diabetes Management. IEEE- 14th Ann. Int. Conf. of the IEEE Eng.in Med. and Biol. Soc., Paris, 1992-CH3207-8, 1238-1239.
3. Szolovits P, Doyle J, Long WJ, Kohane I, Pauker SG. Guardian Angel: Patient-Centered Health Information Systems. Technical Report MIT/LCS/TR-604, Cambridge (MA), 1994.
4. Bellazzi R, Cobelli C, Gomez E and Stefanelli M. The T-IDDM project: Telematic management of Insulin Dependent Diabetes Mellitus. in: *Proceedings of Health Telematics '95*, 1995, M. Bracale, F. Denoth eds., Ischia, 271-276.
5. Bellazzi R, Siviero C, Stefanelli M, De Nicolao G. Adaptive controllers for intelligent monitoring. *Artificial Intelligence in Med.*, 1995: 7: 515-540.
6. Riva A, Bellazzi R. High level control strategies for diabetes therapy. in: *Lecture Notes in Medical Informatics*, 1995, P. Barahona, M. Stefanelli and J. Wyatt eds., Springer Verlag, Berlin, 185-196.
7. Riva A, Bellazzi R and Stefanelli M. Distributed AI Technologies for Patient Management. AAAI '96 Spring Symposium on Artificial Intelligence in Medicine, Stanford (CA), March 25-27, 1996.
8. Deutsch T, Lehmann ED, Carson ER, Roudsari AV, Hopkins KD, Sönksen P. Time series analysis and control of blood glucose levels in diabetic patients. *Comp. Meth. and Programs in Biomed.*, 1994:41:167-182.
9. Frees EW. Data Analysis using Regression Models. Prentice Hall, Englewood Cliffs, NJ, 1996.
10. Shahar Y, Musen MA. Knowledge-Based Temporal Abstraction in Clinical Domains. *Artificial Intelligence in Med.*, to appear.
11. Larizza C, Bernuzzi G, Stefanelli M. A General Framework for Building Patient Monitoring Systems. in: *Lecture Notes in Medical Informatics*, 1995, P. Barahona, M. Stefanelli and J. Wyatt eds., Springer Verlag, Berlin, 91-102.